DOCUMENT RESUME

ED 264 269 TM 850 755

AUTHOR Jones, Douglas H.
TITLE Tools of Robustness for Item Response Theory.
PUB DATE Jul 82
NOTE 12p.; In: Item Response Theory and Computerized Adaptive Testing Conference Proceedings (Wayzata, MN, July 27-30, 1982) (TM 850 744).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Estimation (Mathematics); Functions (Mathematics); *Goodness of Fit; Graphs; *Latent Trait Theory; Mathematical Models; *Maximum Likelihood Statistics; *Measurement Techniques; Psychometrics
IDENTIFIERS *Ability Parameters; *Robustness

ABSTRACT

This paper briefly demonstrates a few of the possibilities of a systematic application of robustness theory, concentrating on the estimation of ability when the true item response model does and does not fit the data. The definition of the maximum likelihood estimator (MLE) of ability is briefly reviewed. After introducing the notion of outlier-pair observation, the effect of the model's lack of fit to the data is demonstrated. Then, a new estimator not as influenced as the MLE by outlier pairs is introduced. Based on the motivation for the new estimator, an entire class of new estimators is provided as a vehicle for understanding two notions of robustness theory: (1) the influence function, which compactly yields the majority of the behavioral results about an estimator; and (2) asymptotic relative efficiency (ARE), a concept in classical statistics and used in item response theory (IRT). Although the notion of ARE has had limited application in IRT, it is a useful alternative to expensive monte carlo simulations of the efficiencies of estimators. One of the drawbacks to determining the ARE of an estimator is the analytic derivation of the asymptotic; however, an expansion using the influence function easily yields the needed asymptotic variances. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Tools of Robustness for Item Response Theory

Douglas H. Jones
Educational Testing Service

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☒ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED264269

The extension of robustness theory (Huber, 1981) to item response theory (IRT) holds great potential for testing. One primary benefit will be increased understanding of the behavior of the usual estimators of item and ability parameters when the assumed model does not hold. Another benefit will be new estimators resistant to departures from a chosen model.

The purpose of this paper is to briefly demonstrate just a few of the possibilities of a systematic application of robustness theory, concentrating on the estimation of ability when the true item response model does and does not fit the data. First, the definition of the maximum likelihood estimator (MLE) of ability will be briefly reviewed. Secondly, after introducing the notion of outlier-pair observation, the effect of the model's lack of fit to the data will be demonstrated. This will provide motivation for the introduction of a new estimator not as influenced as the MLE by outlier pairs. Based on the motivation for the new estimator, an entire class of new estimators, labeled h estimators, will then be introduced.

The h estimators will provide a vehicle for understanding two important notions of robustness theory: (1) the influence function, which compactly yields the majority of the behavioral results about an estimator and (2) asymptotic relative efficiency (ARE), a well-known concept in classical statistics and one which is of use in IRT.

Although the notion of ARE has had limited application in IRT, it is a useful alternative to expensive monte carlo simulations of the efficiencies of estimators. A goal of this paper is thus to further its utility. One of the drawbacks to determining the ARE of an estimator is the analytic derivation of the asymptotic variance; however, an expansion using the influence function easily yields the needed asymptotic variances.

Notation and Form of the Equation Defining the MLE

As usual, denote the response vector in an n-item test as $\underset{\sim}{x} = (x_1, \ldots, x_n)$; the value of the ability parameter will be denoted by θ. For a given θ, the probabilities of correct and incorrect response to the ith item are $P_i(\theta)$ and $Q_i(\theta)$, where $P_i(\theta) + Q_i(\theta) = 1$.

Under the assumptions of local independence, the likelihood function is

TM 850 755

$$L(\theta) = \prod [P_i(\theta)]^{x_i} [Q_i(\theta)]^{1-x_i} . \quad [1]$$

After taking logarithms and the first derivatives with respect to θ, assuming derivatives exist, the normal equation defining the MLE is

$$\sum \frac{P_i'}{P_i Q_i}(\theta)[x_i - P_i(\theta)] = 0 . \quad [2]$$

For the 2-parameter logistic model,

$$[P_i(\theta)/Q_i(\theta)] = a_i(\theta - b_i) \quad [3]$$

and

$$P_i'/(P_i Q_i) = a_i , \quad [4]$$

where a_i is the discrimination parameter and b_i is the difficulty parameter of the ith item. The normal equation takes the form

$$\sum a_i[x_i - P_i(\theta)] = 0 . \quad [5]$$

Example of the Effect of an Outlier

In order to modify the MLE to cope with outliers, there must be a well-defined notion of outliers and an understanding of their effects. For example, let $n = 10$, $b_i = -0.8$ to 1.0 by steps of 0.2, and $a_i = 1$. Table 1 shows two 10-item response vectors (one without and one with an outlier at the 10th response) and their MLEs, where the items have been ordered in increasing difficulty. Apparently, the correct response to the most difficult item unwarrantedly pulls the value of the MLE by almost one-quarter of the range of the difficulty parameters; thus, the single outlier has a substantial effect on the MLE.

Table 1
MLEs for Two 10-Item Response Vectors

$\underline{x}$	MLE
1111100000	.10
1111100001	.58

Further inspection of the shape of the likelihood function for the second response vector in Table 1 would reveal that it is very flat in the region of the extrema; this is one way the likelihood has of warning that the value of the MLE should be treated very cautiously. However, there cannot be misgivings about the value of an estimator when immediate action is called for based on that value, especially for the computerized adaptive testing environment, which

requires selection of future items based on current estimates of an examinee's ability. Therefore, estimators that can cope with outliers must be developed by reducing their effect to give estimates that are not too perturbed from estimates yielded by outlier free response vectors.

Definition of an Outlier Pair

The nature of an outlier, as demonstrated in the above example, is revealed only in reference to the value of the ability and the item characteristic function. Thus, the definition of an outlier requires the inclusion of the item characteristic function. Define $[x_i, P_i(\theta)]$ to be an outlier pair whenever $x_i = 1$ but $P_i(\theta) \simeq 0$ or whenever $x_i = 0$ but $P_i(\theta) \simeq 1$. Note that whenever $[x_i, P_i(\theta)]$ is an outlier, $P_i(\theta)Q_i(\theta) \simeq 0$. The next definition of a new estimator will take advantage of this fact.

A New Estimator

Define an estimator $\hat{\theta}$ by the equation

$$\Sigma a_i[x_i - P_i(\hat{\theta})]\ [P_i(\hat{\theta})\ Q_i(\hat{\theta})] = 0\ . \qquad [6]$$

To understand this modification of the normal equation for the MLE, the example will be continued. Table 2 shows the two θ estimates for the same item response vectors. Thus, the PQ factor weights down the outlier by taking account of the reference value of $P_i(\theta)$ for the 10th item. This idea is continued in the following section.

Table 2
MLE and Modified
Estimate for Two
10-Item Response Vectors

$\underset{\sim}{x}$	MLE	$\hat{\theta}$
1111100000	.10	.10
1111100001	.58	.22

The h Estimators

An entire class of estimators is available for each value of h in the following equation:

$$\Sigma a_i[x_i - P_i(\theta)]\ [P_i(\theta)\ Q_i(\theta)]^{h-1} = 0 \qquad [7]$$

where a_i are real and arbitrary, but given, constants and $h \geq 1$. For $h = 1$, the equation yields the MLE. Once again, the example is continued in Table 3 to observe the effect of this modification.

Table 3
$\underline{h}$ Estimates for
Two 10-Item Response Vectors

$\underline{x}$	h				
	1(MLE)	1.5	2	3	4
1111100000	.10	.10	.10	.10	.10
1111100001	.58	.41	.22	.20	.19

The $\underline{h}$ estimators have the following properties:

1. These estimators do not modify the MLE when the outlier is absent, since whenever $x_i - P_i(\theta)$ is small, the factor $P_i(\theta) \times Q_i(\theta)$ has little effect on the estimator.
2. This class of estimators is also available for any item response model, not just the 2-parameter logistic model. The normal equations have the form

$$\Sigma \frac{P_i'}{P_i Q_i}(\theta)[x_i - P_i(\theta)]\ [P_i Q_i(\theta)]^{h-1} = 0 \ . \qquad [8]$$

3. The asymptotic variance of the estimator is found by differentiating the normal equation, a step similar to taking the second derivative of the log-likelihood function, but this step is not justified by the analogy. The known asymptotic behavior of the MLE depends on the individual observations being <u>independently</u> and <u>identically distributed</u>, a condition that does not generally hold in IRT. The success of extending robustness to IRT is based on extending this condition to IRT in a natural way. This has been accomplished in Jones (1982), where the stochastic structure of IRT has been imbedded in a framework that yields independently and identically distributed observations. Although a simple concept, the precise details of the imbedding will not be explained here; however, some of the tools, such as the limit theorems for the empirical distribution function are briefly discussed as they are used in the following sections.

The Influence Function

Before defining what is meant by the term influence function, an explanation of why they are needed is in order. First, view the ability $\underline{h}$ estimator as a function of $\underline{n}$ observations, where each observation is a triplet denoted by z_i, consisting of (1) the response x_i; (2) the item characteristic function $P_i(t)$, $-\infty < t < \infty$; and (3) the constant a_i. The normal equation, Equation 7, for the $\underline{h}$ estimator is a sum whose summands are specific combinations of the components of the triplet observations. If F_n denotes the empirical distribution function of the $\underline{n}$ triplet observations (z_i)--that is, F_n gives point mass 1/n to each z_i and

zero mass elsewhere--then the normal equation, Equation 7, may be written as the following Lebesque-Stieltjes integral:

$$\int g(z,\theta)\, dF_n = 0 \qquad [9]$$

where

$$g(z,t) = a[x - P(t)][PQ(t)]^{h-1} . \qquad [10]$$

By introducing the above integral, it has been demonstrated that the ability estimator θ is a function of F_n, $\theta(F_n)$, defined implicitly in Equation 9. Furthermore, the integral equation will accept any distribution function, not just F_n. Therefore, the domain of the estimator θ has also been extended to any F; thus $\theta = \theta(F)$. The reason this has been done is to facilitate the introduction of the influence function and eventually to derive its form.

The goal of these developments is to derive a useful approximation to the change of the value of an estimator attributable to the introduction of one more observation. If the observation also happens to be an outlier, then it can be determined how resistant an estimator is to the outlier. The change in the value of an estimator can be easily defined. Let F_n and F_{n+1} be the empirical distributions for n observations and the same n observations plus one more observation. The change in the value of the estimator due to the additional observation is simply

$$\theta(F_{n+1}) - \theta(F_n) . \qquad [11]$$

The two distribution functions are close enough to allow an approximation to Equation 11 by the mean value theorem. Let $\delta(z,z_{n+1})$ be the point mass at z_{n+1}, then

$$F_{n+1}(z) = (1-s)F_n(z) + s\ \delta(z,z_{n+1}) , \qquad [12]$$

where $s = 1/(n+1)$. Since $s \to 0$ for increasing sample size, the derivative useful here has been shown in Jones (1982) to be the derivative of $\theta[(1-s)F_n + s\delta_{n+1}]$ with respect to s and evaluated at $s = 0$. Denoting this derivative by $\dot{\theta}(F_n; z_{n+1})$ gives the needed approximation to Equation 11:

$$\theta(F_{n+1}) - \theta(F_n) \simeq \dot{\theta}(F_n; z_{n+1})/(n+1) . \qquad [13]$$

When viewed as a function of z_{n+1}, the derivative in Equation 13 is called the influence function of the estimator. Jones (1982) has shown the influence function of the estimators, defined by normal equations of the form Equation 9, to be

$$\dot{\theta}(F_n;z) = g(z,\theta) \Big/ \frac{d}{dt} \int g(z,t)dF_n \Big|_{t=\theta} ; \qquad [14]$$

and for the h estimator, this has the form for each observation z_i

$$\dot{\theta}(F_n;z_i) = \frac{[x_i - P_i(\theta)]\ w(\theta;z_i,h)}{\frac{d}{dt} \int [x-P(t)]\ w(t;Z,h)dF \Big|_{t=\theta}} \qquad [15]$$

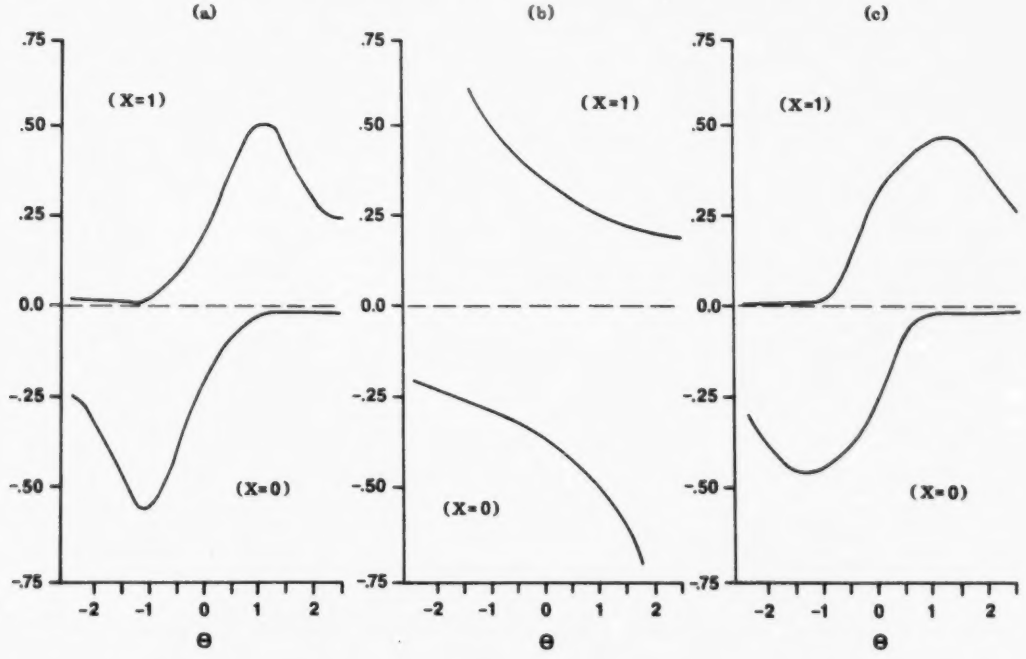
where

$$w(t;z_i,h) = a_i[P_i(t)\ Q_i(t)]^{h-1} \ . \qquad [16]$$

Graphs of Influence Functions: Influence Curves

Since the influence function is a function in $\underline{z}$, a triplet variable, it cannot simply be represented by a single graph of the mapping of the real line. However, if viewed as a function of θ when x_i is either 0 or 1, then it can be conveniently represented by two graphs of the mapping of the ability interval. Figures 1a and 1b contain the graphs of the influence functions (which will be called the influence curves) of two $\underline{h}$ estimators for h = 4 and for h = 1. When h = 1, the $\underline{h}$ estimator is just the MLE. The top curves in Figure 1 (x = 1) correspond to the influence of a correct response to the most difficult item of 10 items; the bottom curves (x = 0) are for an incorrect response to the easiest of the items.

Figure 1
Influence Curves for $\underline{h}$ Estimators with (a) h = 4,
(b) 2-Parameter MLE, and (c) Biweight Estimator with c = 2



From the figures it can be seen that for both h = 4 and h = 1, when x = 1 (a correct response) and θ is the true ability, the observations would have little effect on the estimator when the value of θ is close to 2. However, when θ is close to -2, the observation would have a very large effect for h = 1 (that is, the MLE) but little effect for an $\underline{h}$ estimator when h = 4. These analytic circumspections correspond precisely to the conclusons indicated by the examples

of the previous section. How much of an effect a correct response has to the most difficult item for a person of low ability can be obtained directly from Figures 1a and 1b. The values of the influence curves are approximately 0 for the h estimator and .80 for the MLE; this latter value is about 40% of the total ability scale. Figure 1c will be discussed below.

Figure 2
Influence Curves for h Estimator with (a) h = 10
and (b) Trimmed 2-Parameter MLE

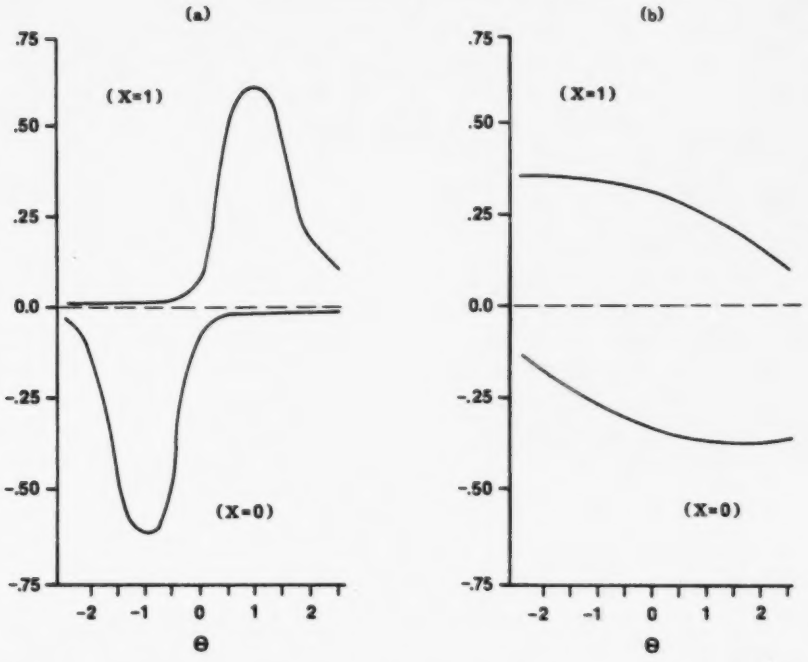


Figure 2a shows the general shape of the influence curves where h is large, e.g., h = 10. The estimator for even larger h tends to treat both correct and incorrect responses equivalently while allowing only responses to items for which the probability of a correct response is close to .5 to affect the estimator. Figure 2b is based on a new estimator discussed below.

The Properties Yielded by Influence Functions

Besides yielding an analysis of the effect of a single observation on the value of an estimator, influence functions also yield the following: (1) new robust estimators, (2) the asymptotic distributions of the estimators, and (3) comparison between competing estimators.

Derivations of New Estimators That Are Robust

For new robust estimators, new estimators can be derived simply by specifying the graph of the influence function. Figure 2b is an example providing a

MLE for correct responses when $\theta > 0$ and for incorrect responses when $\theta < 0$; however, it differs from the MLE for correct responses where $\theta < 0$ by giving less influence to the observation for each value of $\theta < 0$. The estimator behaves similarly for incorrect responses.

Asymptotic Distributions of Estimators

For the asymptotic distributions of estimators, Jones (1982) exploited the influence function approximation to the h estimator in order to derive its limiting distribution. If the items are randomly sampled according to a fixed distribution, then F_n will converge to a limiting distribution F (Jones, 1982). The parameters of the limiting distribution of the sequence of the estimators depend on the F, as the limiting distribution is normal with mean $\theta(F)$ and variance

$$V(F,\theta;h) = \frac{\int [x-P(\theta)]^2 \, w^2(\theta;z,h)dF}{\left[\frac{d}{dt}\int [x-P(t)] \, w(t;Z,h)dF\Big|_{t=\theta}\right]^2} \qquad [17]$$

Comparisons Between Competing Estimators

Finally, comparison between competing estimators includes (1) computing the asymptotic relative efficiency and (2) visual comparison of the influence graphs.

Asymptotic relative efficiency. The usefulness of the asymptotic relative efficiency will be demonstrated for the h estimators relative to the MLE. The efficiency is defined for unbiased estimators as the ratio of the asymptotic variances

$$Eff = V(F,\theta;h)/V(F,\theta;1) \ , \qquad [18]$$

and for biased estimators as the ratio of the mean squared errors (the formulas are given in Jones, 1982).

The model under which the estimators are compared correspond to the model in the example of the previous sections. The value of the efficiencies are shown in Table 4. The greatest loss of efficiency by the h estimators is only one item in 10 at h = 5.

Whereas the h estimator is always less efficient than the MLE under the assumed model, the opposite situation can occur under an alternative model. Let the alternative model be generated by perturbing the item response curves by slightly altering the item difficulty parameters; that is, for each item let the item response curve be equal to

$$P_i^*(t) = P_i(t-.1) \ . \qquad [19]$$

The resulting efficiencies are also shown in Table 4. Thus, as Table 4

shows, the efficiency of the h estimator can be far superior to the MLE under a model different from the assumed one.

Table 4
Values of Efficiency for Five
Levels of h under Assumed
and Alternative Models

	h				
Model	1.5	2.0	3.0	4.0	5.0
Assumed	.99	.98	.95	.92	.90
Alternative	2.24	4.83	19.15	51.46	84.62

Comparison among the biweight, Bayes, 3-parameter MLE, and the h estimators. Visual comparison of influence graphs of estimators is highly enlightening. The influence graphs of three estimators--the Bock-Mislevy biweight estimator (Mislevy & Bock, 1981), the Bayes posterior mode, and the MLE for the 3-parameter logistic model--will be illustrated.

The influence function of the biweight and MLE for the 3-parameter logistic model is the same as Equation 15 except that the form of the w function for the biweight is

$$w(t;z_i) = (1-t^2)^2 \quad \text{for } |t| \leq c$$
$$= 0 \quad \text{otherwise} . \qquad [20]$$

and for the 3-parameter logistic is

$$w(t;z_i) = \frac{R_i(t)}{c_i + (1-c_i)\,R_i(t)} , \qquad [21]$$

where $R_i(t)$ is a 2-parameter logistic curve and $P_i(t) = c_i + (1-c_i)\,R_i(t)$. The influence function of the Bayes posterior mode with prior $\pi(t)$ is

$$\dot{\theta}(F;z_i) = \frac{u(\theta,z_i)[x_i - P_i(\theta)]}{\frac{d}{dt}\int u(t,z)[x - P(t)]dF\Big|_{t=\theta} + \psi'(\theta)} \qquad [22]$$

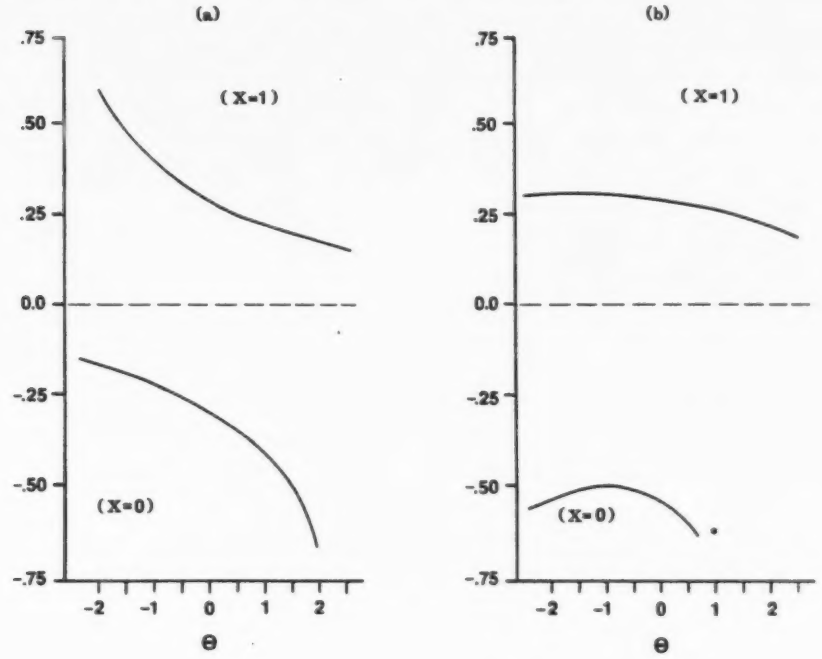
where

$$\psi(t) = \pi'(t)/\pi(t) \qquad [23]$$

and

$$u(t,z) = P'(t)/PQ(t) . \qquad [24]$$

Figure 3
Influence Curves for Posterior Mode Estimator of (a) 2-Parameter Logistic Likelihood and Normal Prior with Mean 0 and Standard Deviation 5, and (b) 3-Parameter MLE with Guessing Parameter .2



Figures 1c, 3a, and 3b illustrate the influence curves of all three estimators. It can be seen that the biweight estimator with c = 2 behaves similarly to the h estimator except that the biweight allows influence for a smaller interval of ability and, for the interval, the influence it allows is slightly larger than the h estimator. The Bayes posterior modal estimator is very similar to the 2-parameter MLE except that it uniformly downweights observations; it is surprising that there is any downweighting at all, since the standard deviation of the prior is 5 for this example. The 3-parameter MLE estimator has the most surprising influence curve of all. For correct responses, it is similar to the 2-parameter MLE, but it upweights observations with ability near the upper extreme; similarly, for incorrect responses it inflates the influence over the 2-parameter MLE, especially for large abilities. This observation is suprising because the 3-parameter MLE has been maligned as being unreliable, since the item characteristic parameters are difficult to estimate accurately. The influence curve shows, however, that the 3-parameter MLE possesses intrinsic and probably accidental robustness against outliers for correct responses, although at the expense of some undesirable sensitivity to incorrect responses. The asterisk in Figure 3b indicates that the influence curve continues off the graph, ending at -1.75 for $\theta = 2$.

REFERENCES

Huber, P. J. Robust statistics. New York: Wiley, 1981.

Jones, D. H. Redescending M-type estimators of latent ability (ETS RB-82-30). Princeton NJ: Educational Testing Service, 1982.

Mislevy, R. J., & Bock, R. D. Biweight estimates of latent ability. Unpublished manuscript, 1981.